code|cademy

Capstone: Attribution

Learn SQL from Scratch
Awongiwe Ntaote
29 August 2018

Table of Contents

1. Get familiar with CoolTShirts
2. What is the user journey?
3. Optimize the campaign budget

1. Get familiar with CoolTShirts

Data Samples

user_id	first_touch_at	utm_source	utm_campaign	page_name
10006	2018-01-24 03:12:16	nytimes	getting-to-know-cool-tshirts	1 - landing_page
10030	2018-01-25 20:32:02	buzzfeed	ten-crazy-cool-tshirts-facts	1 - landing_page
10045	2018-01-05 18:31:17	nytimes	getting-to-know-cool-tshirts	1 - landing_page
10048	2018-01-16 04:17:46	medium	interview-with-cool-tshirts-founder	1 - landing_page
10069	2018-01-02 23:14:01	buzzfeed	ten-crazy-cool-tshirts-facts	1 - landing_page
10162	2018-01-29 21:37:10	nytimes	getting-to-know-cool-tshirts	1 - landing_page
10177	2018-01-24 07:10:33	nytimes	getting-to-know-cool-tshirts	1 - landing_page
10254	2018-01-23 22:27:18	medium	interview-with-cool-tshirts-founder	1 - landing_page
10329	2018-01-18 05:27:25	medium	interview-with-cool-tshirts-founder	1 - landing_page
10354	2018-01-19 10:57:29	nytimes	getting-to-know-cool-tshirts	1 - landing_page

first_touch query

```sql
WITH first_touch AS (
    SELECT user_id,
        MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT ft.user_id,
    ft.first_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    pv.page_name
FROM first_touch ft
JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp;
```

last_touch query

```sql
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    pv.page_name
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp;
```

user_id	last_touch_at	utm_source	utm_campaign	page_name
10006	2018-01-25 23:10:16	email	weekly-newsletter	3 - checkout
10030	2018-01-28 13:38:02	email	retargetting-campaign	4 - purchase
10045	2018-01-09 03:05:17	facebook	retargetting-ad	3 - checkout
10048	2018-01-19 00:00:46	email	retargetting-campaign	3 - checkout
10069	2018-01-04 08:13:01	facebook	retargetting-ad	4 - purchase
10162	2018-02-01 04:26:10	email	weekly-newsletter	4 - purchase
10177	2018-01-24 11:58:33	nytimes	getting-to-know-cool-tshirts	2 - shopping_cart
10254	2018-01-25 09:12:18	google	paid-search	3 - checkout
10329	2018-01-22 16:35:25	email	retargetting-campaign	4 - purchase
10354	2018-01-22 02:00:29	google	paid-search	4 - purchase

1.1 Question 1

How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?

Use three queries:

- one for the number of distinct campaigns,
- one for the number of distinct sources,
- one to find how they are related.

Campaigns	
8	
Sources	
6	
Campaigns	**Sources**
getting-to-know-cool-tshirts	nytimes
weekly-newsletter	email
ten-crazy-cool-tshirts-facts	buzzfeed
retargetting-campaign	email
retargetting-ad	facebook
interview-with-cool-tshirts-founder	medium
paid-search	google
cool-tshirts-search	google

```
SELECT COUNT(DISTINCT utm_campaign) AS 'Campaigns'
FROM page_visits;


SELECT COUNT(DISTINCT utm_source) AS 'Sources'
FROM page_visits;


SELECT DISTINCT utm_campaign AS 'Campaigns',
            utm_source AS 'Sources'
FROM page_visits;
```

There are 8 campaigns and 6 sources used by CoolTShirts.

Each source used the these campaigns:

Nytimes: getting-to-know-cool-tshirts
Email: weekly-newsletter & retargeting-campaign
Buzzfeed: ten-crazy-cool-tshirts-facts
Facebook: retargeting-ad
Medium: interview-with-cool-tshirts-founder
Google: paid-search & cool-tshirts-search

1.2 Question 2

What pages are on the CoolTShirts website?

Find the distinct values of the page_name column.

page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

page_name	COUNT(DISTINCT page_name)
3 - checkout	4

```
SELECT page_name
FROM page_visits
GROUP BY 1;


SELECT page_name, COUNT(DISTINCT page_name)
FROM page_visits;
```

The pages on the CoolTShirts website are:
- **1 – landing_page**
- **2 – shopping_cart**
- **3 – checkout**
- **4 – purchase**

The distinct values of the page_name column is 4.

2. What is the user journey?

2.1 Question 3

How many first touches is each campaign responsible for?

Sources	Campaigns	Total First Touches
medium	interview-with-cool-tshirts-founder	622
nytimes	getting-to-know-cool-tshirts	612
buzzfeed	ten-crazy-cool-tshirts-facts	576
google	cool-tshirts-search	169

The total number of first touches for each campaign were as follows:

- **interview-with-cool-tshirts-founder - 622**
- **getting-to-know-cool-tshirts - 612**
- **ten-crazy-cool-tshirts-facts - 576**
- **cool-tshirts-search - 169**

```sql
WITH first_touch AS (
SELECT user_id,
        MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id),
ft_attr AS (
  SELECT ft.user_id,
         ft.first_touch_at,
         pv.utm_source,
         pv.utm_campaign
  FROM first_touch ft
  JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp
)
SELECT ft_attr.utm_source AS 'Sources',
       ft_attr.utm_campaign AS 'Campaigns',
    COUNT (*) AS 'Total First Touches'
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2.2 Question 4

How many last touches is each campaign responsible for?

Sources	Campaigns	Total Last Touches
email	weekly-newsletter	447
facebook	retargetting-ad	443
email	retargetting-campaign	245
nytimes	getting-to-know-cool-tshirts	232
buzzfeed	ten-crazy-cool-tshirts-facts	190
medium	interview-with-cool-tshirts-founder	184
google	paid-search	178
google	cool-tshirts-search	60

The total number of last touches for each campaign were as follows:
- **weekly-newsletter – 447**
- **retargeting-ad – 443**
- **retargeting-campaign – 245**
- **getting-to-know-cool-tshirts - 232**
- **ten-crazy-cool-tshirts-facts - 190**
- **interview-with-cool-tshirts-founder – 184**
- **paid-search - 178**
- **cool-tshirts-search - 60**

```sql
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id),
lt_attr AS (
  SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign,
        pv.page_name
  FROM last_touch lt
  JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source  AS 'Sources',
        lt_attr.utm_campaign AS 'Campaigns',
        COUNT(*) AS 'Total Last Touches'
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2.3 Question 5

How many visitors make a purchase?

Count the distinct users who visited the page named `4 - purchase`.

361 visitors made a purchase on CoolTShirts

```sql
SELECT COUNT(DISTINCT user_id) AS 'Total Purchases'
FROM page_visits
WHERE page_name = '4 - purchase';
```

Total Purchases
361

2.4 Question 6

How many last touches on the purchase page is each campaign responsible for?

This query will look similar to your last-touch query, but with an additional WHERE clause.

Sources	Campaigns	Total Purchases
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

Each campaign is responsible for the following purchases:
*   **weekly-newsletter – 115**
*   **retargeting-ad – 113**
*   **retargeting-campaign – 54**
*   **paid-search – 52**
*   **ten-crazy-cool-tshirts-facts - 9**
*   **getting-to-know-cool-tshirts - 9**
*   **interview-with-cool-tshirts-founder – 7**
*   **cool-tshirts-search - 2**

```sql
WITH last_touch AS (
    SELECT user_id,
           MAX(timestamp) as last_touch_at
    FROM page_visits
    WHERE page_name = '4 - purchase'
    GROUP BY user_id),
lt_attr AS (
  SELECT lt.user_id,
         lt.last_touch_at,
         pv.utm_source,
         pv.utm_campaign,
         pv.page_name
  FROM last_touch lt
  JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source AS 'Sources',
       lt_attr.utm_campaign AS 'Campaigns',
       COUNT(*) AS 'Total Purchases'
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

3. Optimize the campaign budget

3. Optimize the campaign budget

CoolTShirts can re-invest in 5 campaigns.

Which should they pick and why?

The 5 campaigns that CoolTShirts can re-invest in are:

- **weekly-newsletter**
- **retargeting-ad**
- **retargeting-campaign**
- **paid-search**
- **getting-to-know-cool-tshirts**

weekly-newsletter, retargeting-ad, retargeting-campaign and paid-search have the highest number of purchases. Which means that there was a return on investment, so it would be a good idea for CoolTShirts to re-invest in these 4 campaigns.

The 5th campaign they can choose is getting-to-know-cool-tshirts because is has a high first touch of 612 and a last touch of 232. This campaign can be re-invesed in because it has the potential to improve. CoolTShirts can use it for brand awareness and getting more people to visit their website with aim of getting visitors to make a purchase.

Sources	Campaigns	Total First Touches
medium	interview-with-cool-tshirts-founder	622
nytimes	getting-to-know-cool-tshirts	612
buzzfeed	ten-crazy-cool-tshirts-facts	576
google	cool-tshirts-search	169

Sources	Campaigns	Total Last Touches
email	weekly-newsletter	447
facebook	retargetting-ad	443
email	retargetting-campaign	245
nytimes	getting-to-know-cool-tshirts	232
buzzfeed	ten-crazy-cool-tshirts-facts	190
medium	interview-with-cool-tshirts-founder	184
google	paid-search	178
google	cool-tshirts-search	60

Sources	Campaigns	Total Purchases
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2